

Standarddeutsch messen? Frequenz und Varianz negativ-konditionaler Konnektoren

Marek Konopka, Ulrich H. Waßner
unter Mitwirkung von Sandra Hansen¹
Institut für Deutsche Sprache (IDS), Mannheim
konopka@ids-mannheim.de, wassner@ids-mannheim.de

Measuring Standard German? Frequency and Variance of Negative-Conditional Connectors

ABSTRACT: Among the German negative-conditional connectors in the range of *consequens* markers there are the prototypical cases *sonst* and *ansonsten*. Morphological alternatives (*sonsten* and *ansonst*) are rarely mentioned in contemporary grammars and dictionaries but they actually occur with considerable frequency. The four connectors are used in two functions: as a conjunctive adverb which can occupy various positions within the sentence or as a specific kind of subordinating conjunction (*Postponierer*). The large IDS corpora allow us to reveal specific distributions of the lexemes and of their different ways of use. Comparing the frequencies and the distributions can indicate to which extent the phenomena are part of the standard language. The paper will report on the results and demonstrate how the findings can be deduced from the corpora. It will draw conclusions for assessing the acceptability of the variants and the extent to which they can be considered standard language additionally testing statistical instruments to visualise and calculate the variance of phenomena as association plots and DP_{norm} .

KEY WORDS: corpus linguistics, grammar, standard German, connectors, conjunction, adverbials, variation

Jak měřit standardní němčinu? Frekvence a variantnost negativně podmiňovacích konektorů

ABSTRAKT: Prototypickými případy těchto konektorů jsou v němčině slova *sonst* (*jinak*) a *ansonsten* (*jinak*). Jejich morfologické alternativy *sonsten* a *ansonst* se v součas-

¹ Sandra Hansen hat computerlinguistische Aufgaben wie die Extraktion der Daten übernommen und zahlreiche statistische Analysen (vor allem in R) durchgeführt. Wir danken außerdem Saskia Schmadel für die Unterstützung bei mühsamen Korpusrecherchen und -auswertungen, bei der Erstellung von Abbildungen und der Formatierung. Last not least sind wir Anja Konopka für die wertvollen Korrekturen verbunden.

ných gramatikách a slovnících téměř nezmiňují, avšak jejich frekvence v textech není zanedbatelná. Všechny tyto čtyři konektory se užívají ve dvou funkcích: jako spojkové adverbium, jež může ve větě zaujímat různé pozice, nebo jako specifický druh subordinační konjunkce (postponent). Velké korpusy IDS umožňují odhalovat specifickou distribuci těchto lexémů a různých způsobů jejich užívání. Srovnávání frekvence a distribuce může indikovat, do jaké míry jsou tyto jevy součástí standardního jazyka. Tato studie přináší výsledky tohoto srovnávání a ukazuje, jak lze korpusové nálezy interpretovat. Prostřednictvím testování statistických nástrojů se vyslovuje k akceptabilitě variant a míře, v níž mohou náležet standardnímu spisovnému jazyku.

KLÍČOVÁ SLOVA: korpusová lingvistika, gramatika, standardní němčina, konektory, spojky, adverbiale, varianta

0 Einführung

Im Rahmen des Projekts „Korpusgrammatik“ werden am Institut für Deutsche Sprache (IDS) in Mannheim korpuslinguistische Verfahren entwickelt, um die Standardsprachlichkeit grammatischer Konstruktionen zu diagnostizieren und die Grammatik des Standarddeutschen umfassend zu erfassen. Das langfristige Ziel des Projekts ist eine neue, korpuslinguistisch fundierte Grammatik des Deutschen mit einem Schwerpunkt auf der Variation der Standardsprache.² Im vorliegenden Beitrag wird eine kleinere Pilotstudie vorgestellt, in der einerseits die Variation der Konnektoren der *sonst*-Gruppe untersucht wird und andererseits einige Maße und Visualisierungsformen evaluiert werden, mit denen Frequenz und Verteilung im Korpus in Abhängigkeit von den Faktoren Zeit, Raum, Register und syntaktische Funktion erfasst werden können. Die Vorgehensweise umfasst folgende Punkte:

- Allgemeine Informationen zur Methodik der Korpusrecherche und -analyse
- Überblick über Konnektoren der *sonst*-Gruppe: gemeinsame morphologische, syntaktische und semantische Charakteristika
- Untersuchung der Lexemvarianten: Frequenz und Varianz
- Untersuchung der Gebrauchsvarianten: Frequenz und Varianz verschiedener syntaktischer Verwendungen, Gebrauchspräferenzen der Lexeme
- Linguistische und methodologische Folgerungen

1 Allgemeine Informationen zur Methodik der Korpusrecherche und -analyse

Das langfristige Ziel des Projekts ist eine Referenz-Grammatik des Deutschen, die als von der Englisch-Grammatik von Biber et al. (1999; 2006) inspiriert bezeichnet werden kann, deren genauere Ausrichtung und eigenständiger Aufbau aber erst nach Abschluss der Pilotstudien festgelegt werden

² Vgl. <http://www1.ids-mannheim.de/gra/projekte/korpusgrammatik.html> sowie <http://hypermedia.ids-mannheim.de/call/public/korpus.ansicht>; zu einer frühen Beschreibung des Projekts und der allerersten Pilotstudie vgl. Kubczak – Konopka 2008.

sollen. In Letzteren sollen quasi probeweise überschaubare, variationsreiche Ausschnitte der deutschen Grammatik untersucht werden, um nach und nach die Methodologie zur Erreichung des langfristigen Ziels zu erarbeiten. In der vorliegenden Pilotstudie stehen linguistisch gesehen die negativ-konditionalen Konnektoren *sonst*, *ansonsten*, *ansonst* und *sonsten* zur Diskussion und methodologisch gesehen quantitativ-statistische Werkzeuge zur Diagnose und Spezifizierung der Standardsprachlichkeit.

Die Untersuchungen basieren auf Auswertungen von Korpora des IDS, vor allem des *Gesamtkorpus* des Projekts „Korpusgrammatik“³, das aus dem Deutschen Referenzkorpus (DeReKo, Release vom 29. 03. 2011)⁴ hervorgegangen ist und zum Zeitpunkt der Analysen (November 2012) über 4 Mrd. Token bzw. ca. 16 Mio. (geschriebene) Texte umfasste. Die Korpustexte waren mit den Taggern „Machine Phrase Tagger“ der Firma Connexor Oy und „Tree-Tagger“ annotiert⁵ und mit Metadaten angereichert, die die Variationsparameter Zeit, Medium, Register, Staat, Region und inhaltliche Domäne spezifizierten.

Die Untersuchung umfasste folgende Analyseschritte:

- Es wurde die Frequenz des Phänomens im Gesamtkorpus und seine Verteilung auf virtuelle Teilkorpora ermittelt, die entlang der obengenannten Metadaten gebildet werden konnten.
- Die Frequenzen und die Verteilung der Varianten wurden miteinander verglichen und die signifikanten Unterschiede als Indizien dafür betrachtet, wie das Verhältnis der einzelnen Varianten zur Standardsprache einzuschätzen ist.
- Den Indizien folgend wurden punktuell Vergleiche mit anderen Korpora angestellt bzw. die Streuung der Varianten im Korpus und seinen Teilen untersucht.

Dabei wurde – wenn nicht anders vermerkt – bei wortbasierten Suchen und Berechnungen die Groß- bzw. Kleinschreibung, z. B. *Ansonsten* vs. *ansonsten*, nicht beachtet.

Zur Visualisierung der Frequenz- und Verteilungsunterschiede wurden die Assoziationsplots (vgl. Cohen 1980; Friendly 1992; Meyer et al. 2005) verwendet, die mithilfe des *vcd*-Pakets der statistischen Software R⁶ erstellt wurden. Um die Streuung der Phänomene genauer einzuschätzen, wurde eine Reihe von Maßen getestet und schließlich für die Darstellung Gries' normalisierte *Deviation of Proportions* (DP_{norm} , vgl. Gries 2008; 2009)⁷ eingesetzt.

³ Vgl. http://hypermedia.ids-mannheim.de/call/public/korpus.ansicht?v_id=4752.

⁴ Vgl. www.ids-mannheim.de/kl/projekte/korpora/.

⁵ Näheres unter www1.ids-mannheim.de/kl/projekte/korpora/annotationen.html.

⁶ <http://www.r-project.org/>.

⁷ Vergleich und Diskussion der Maße in Bubenhofer (2012).

2 Überblick über die Konnektoren der *sonst*-Gruppe

Unter Konnektoren werden hier Ausdrücke verstanden, die eine semantische Relation zwischen zwei propositionalen Strukturen indizieren.⁸ In traditionelleren Wortartenklassifikationen entsprechen den Konnektoren:

- einerseits koordinierende und subordinierende Konjunktionen (z. B. *und*, *oder* bzw. *wenn*, *weil*) mit Ausnahme der „Komplementierer“ *dass* und *ob*,
- andererseits semantisch „zweistellige“ Konjunkionaladverbien (z. B. *angenommen*, *auch*, *sonst*, *andernfalls*).

Hier wird eine bestimmte, relativ homogene Gruppe von Konnektoren behandelt, vgl.:

- (1) Verträge über den Kauf von Wohnungen müssen eine Baubeschreibung enthalten, *sonst* / *sonsten* / *ansonsten* / *ansonst* sind sie gesetzlich unwirksam.⁹

Die vier Konnektoren können als Varianten betrachtet werden, da sie wesentliche morphologische, syntaktische und semantische Eigenschaften teilen. Die **morphologische** Verwandtschaft beruht vor allem auf der Gleichheit des Wortstammes (*-sonst-*), geht aber auch darüber hinaus (vgl. Tabelle 1).

Tab. 1 – Gemeinsame morphologische Elemente

	-ø	-en
ø-	<i>sonst</i>	<i>sonsten</i>
an-	<i>ansonst</i>	<i>ansonsten</i>

Was die **syntaktischen** Eigenschaften der vier Konnektoren angeht, so haben sie eine nahezu identische Distribution. Sie treten auf als

- nicht-nacherstfähige (*sonst(en)*) bzw. nicht-positionsbeschränkte (*ansonst(en)*) **Adverbkonnektoren** (d. h. sie kommen innerhalb eines Konnektivs vor, und zwar vor allem ins Vorfeld oder ins Mittelfeld integriert, evtl. phrasenintern, vorwiegend mit Adjektivbezug)¹⁰ oder

⁸ Vgl. Pasch et al. (2003 : 331/332 et passim). Es geht also nicht nur um ausformulierte Sätze als Konnekte; auch z. B. die Verwendungen mit Adjektiv(phrasen)bezug werden durch diesen Begriff von Konnektoren abgedeckt, wenn als Denotat des internen Konnektivs ein Sachverhalt in Frage kommt. Dies zeigt sich in der Explizitierbarkeit als (z. B. Relativ-)Satz: *die ansonsten korrekten Deutschen > die Deutschen, die ansonsten korrekt sind*, vgl. ebd., Kapitel B 6 sowie 234, 546–550, 610.

⁹ Originalbeleg: „Auch Verträge über bereits fertiggestellte Wohnungen müssen nämlich eine Baubeschreibung enthalten, **ansonsten** sind sie nach der Rechtsprechung des Bundesgerichtshofes unwirksam.“ (Berliner Zeitung, 10. 2. 2007, S. 1).

¹⁰ Vereinzelt kommen die Adverbkonnektoren *ansonst* und *ansonsten* darüber hinaus in

• **Postponierer** (d. h. sie können ihr internes Konnekt regieren, indem sie in diesem Letztstellung des finiten Verbs bewirken, vgl. Pasch et al. 2003, 351),

aber nicht als Vertreter einer anderen syntaktischen Konnektorenklasse. Man vergleiche folgende Belege für die selteneren Konnektoren *ansonst* und *sonsten*:

• **Adverbkonnektor im Vorfeld**

- (2) Hunde sind an der Leine zu führen! **Ansonst** wird geschossen. (Zürcher Tagesanzeiger, 17. 7. 1996, S. 27)
- (3) Die Senatsverkehrsverwaltung rät, auf öffentliche Verkehrsmittel umzusteigen. **Sonsten** drohe Dauerstau. (Berliner Zeitung, 28. 8. 1999, S. 17)

• **Adverbkonnektor im Mittelfeld**

- (4) [A:] „Es ist eine Frechheit, daß ausgerechnet in der Haupturlaubszeit die Bürgerbegutachtung stattfindet.“ [B:] „Wir könnten **ansonst** die Fristen nicht einhalten.“ (Neue Kronen-Zeitung, 4. 8. 1995, S. 24)
- (5) Mit der DEL hat die letzte der sieben europäischen Top-Ligen ihre Teilnahme bekanntgegeben, als Nachrücker hätte **sonsten** Weißrussland bereit gestanden. (dpa, 9. 4. 2008, o. S.)

• **Adverbkonnektor mit Adjektivphrasenbezug**

- (6) Das Image der **ansonst** so korrekten Deutschen ist beschädigt. (Nürnberger Nachrichten, 29. 1. 2005, o. S.)
- (7) Über die Neigung und Abneigungen, die Stärken und Schwächen seiner goldenen Lieblinge verrät der **sonsten** gesprächige Wirtschaftsjournalist kein Wort. (Die Welt, 31. 10. 1974, S. 21)

• **Postponierer**

- (8) Der Chauffeur hatte noch die Geistesgegenwart gehabt, den Motor abzustellen und die Bremsen zu ziehen, **ansonst** wohl sämtliche Insassen verbrannt wären. (St. Galler Tagblatt, 21. 2. 2000, o. S.)
- (9) „Geboten ist aber, den 2003 begonnenen Weg konsequent fort zu setzen, **sonsten** die Haushaltsprobleme sich noch weiter verschlimmern“, so der Beschlussvorschlag für die Stadträte. (Rhein-Zeitung, 25. 2. 2004, o. S.)

Ähnlich universell ist auch das viel häufigere *ansonsten*. Selbst für die heute etwas ungewöhnliche Postponierer-Verwendung finden sich im DeReKo Belege, z. B.:

der Nullposition vor dem zweiten Konnekt vor, z. B.: „Wer behauptet, den Wagen schon wieder verkauft zu haben, muß die Adresse des Käufers vorweisen. **Ansonsten** – den letzten beißen die Hunde.“ (Frankfurter Rundschau, 16. 4. 1998, S. 1), sowie in der sog. „Nacherstposition“, z. B.: Nur wegen der möglichen Brandgefahr hat der Verein interveniert. Die Aktivität **ansonsten** hat der Verein begrüßt [...]. (Frankfurter Rundschau, 9. 5. 1997, S. 2).

- (10) Es geht doch nicht an, dass ich als Hotelier gut beleumundete, mündige Menschen und Gäste um 24 Uhr ins Bett schicken muss, **ansonsten** ich mich strafbar mache. (St. Galler Tagblatt, 25. 2. 1998, o. S.)

Alle oben genannten Varianten der adverbialen Verwendung sind bekanntlich auch bei dem häufigsten Konnektor *sonst* sehr verbreitet. Allerdings lässt sich für *sonst* anhand gegenwartssprachlicher Texte keine Postponierer-Verwendung belegen, was die einzige Einschränkung der syntaktische Funktionalität bei einem der Konnektoren darstellt.

Semantisch gesehen gehören die vier Konnektoren mit ihrer zentralen Bedeutung gemäß der Klassifikation des zweiten Teils des *Handbuchs der deutschen Konnektoren* (Breindl et al., im Ersch.) den negativ-konditionalen Konnektoren an und darin der Untergruppe der Konsequens-Marker.¹¹ Die negativ-konditionalen Konnektoren wiederum bilden eine von mehreren Klassen der konditional basierten Konnektoren, vgl. folgende Gesamtübersicht:

- a Temporale Konnektoren (z. B.: *während, als, sooft, solange, seit, bis*)
- b Additiv basierte Konnektoren (z. B.: *und, weder ... noch, aber, wobei*)
- c Alternativ basierte (disjunktive) Konnektoren (z. B.: *oder, bzw.*)
- d Konditional basierte Konnektoren
 - d.a konditionale Konnektoren (z. B.: *wenn, falls, angenommen*)
 - d.b kausale Konnektoren (z. B.: *da, weil, denn*)
 - d.c konzessive Konnektoren (z. B.: *obwohl, trotzdem, auch/selbst wenn*)
 - d.d irrelevanzkonditionale Konnektoren (z. B.: *ob ... ob, sowieso, jedenfalls*)
 - d.e finale und instrumentale Konnektoren (z. B.: *damit, indem, wodurch*)
 - d.f negativ-konditionale Konnektoren
 - d.f.a Antezedens-Marker (z. B.: *aufßer, es sei denn*)
 - d.f.b Konsequens-Marker** (z. B.: *sonst(en), ansonst(en), ander(e)n-falls*)
- e Metakommunikative Konnektoren (z. B.: *beispielsweise, besser gesagt, d. h.*)

Alle vier Konnektoren können die Bedeutung ‚im anderen Falle‘ haben, was sich formelhaft etwa so ausbuchstabiert:

$P, (an)sonst(en) \ q = \text{[grundsätzlich } p, \text{ aber] } \underline{\text{wenn}} \text{ nicht } p, \underline{\text{dann}} \ q'.$

In dieser Umschreibung wird sichtbar, warum die Klasse als negativ-konditionale bezeichnet wird. Das externe Konnekt *p* des Konnektors, das Antezedens in der unterliegenden Konditionalrelation, wird – vom Konnektor induziert – negiert. Das interne Konnekt *q* des Konnektors ist das Konsequens-Argument in der zugrunde liegenden konditionalen Relation, weswegen diese Konnektoren als Konsequens-Marker bezeichnet werden.¹² Vgl.:

¹¹ Zu dieser Unterklasse gehören ansonsten nur noch *ander(e)nfalls* und *widrigenfalls*, die ebenfalls als nicht-nacherstfähige Adverbkonnektoren oder Postponierer auftreten.

¹² Aussagenlogisch ist die negativ-konditionale Relation, interpretiert als hinreichende Bedingung mit negiertem Antezedens, äquivalent zur inklusiven Disjunktion; als

- (11) Tausend Rupien lassen den Beamten über die Mängel hinwegsehen – **wenn** er milde gestimmt ist, ansonsten kostet die Nachsicht mehr. (Zürcher Tagesanzeiger, 10. 10. 1998, S. 5) – D. h. **wenn** er **nicht** milde gestimmt ist, kostet die Nachsicht mehr'

Dieser Bedeutung können auch die Belege (1)–(5) sowie (8)–(10) weiter oben zugeordnet werden. Die vier Konnektoren haben auch prinzipiell das weitere Bedeutungsspektrum gemeinsam, wie man etwa an den in Wörterbüchern angegebenen Quasi-Synonymen sehen kann. Zu dem Bedeutungsspektrum gehören

- neben der zentralen, klassenspezifischen, negativ-konditionalen Bedeutung **‚wenn aber das nicht, dann‘**
- vor allem die Bedeutung **‚unter anderen Umständen / in anderen Hinsichten‘** (vgl. Belege (6) und (7) weiter oben) mit speziellen temporalen (‚zu anderen Zeiten‘) und lokalen (‚an anderen Orten‘) Lesarten und
- die additiv basierte Bedeutung **‚außerdem / des Weiteren‘**.

Folgende zwei Belege sollen die beiden letzteren Bedeutungen illustrieren:

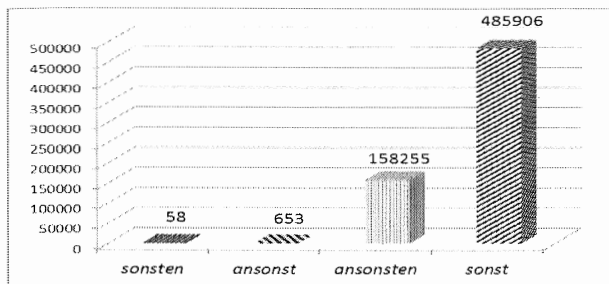
- (12) Bei Heimatfesten oder zu historischen Anlässen wird noch einmal „angestocht“. **Sonsten** bleibt der Ofen kalt – ‚unter anderen Umständen / zu anderen Zeiten‘ (Rhein-Zeitung, 24. 6. 1999)
- (13) Davon müssen alle laufenden Kosten für Miete, Lebensmittel, Sprit fürs Auto und was Europäer in Afrika **sonst** noch fürs tägliche Leben erhalten, gedeckt werden – ‚außerdem / des Weiteren‘ (Rhein-Zeitung, 16. 01. 2006, „Humanitäre Hilfe aus der Luft“)

Die semantischen und syntaktischen Möglichkeiten der vier Konnektoren sind damit weitgehend deckungsgleich. Dennoch gibt es deutliche Unterschiede in ihrer Akzeptabilität, was schon der Vergleich ihrer absoluten Frequenzen im DeReKo vermuten lässt¹³ (vgl. Abbildung 1).

notwendige zur Sheffer-Funktion ‚nicht beide‘; und als notwendige und hinreichende zur exklusiven Disjunktion. Das dürfte der Grund dafür sein, dass einzelne Vertreter dieser Klasse von Konnektoren – vor allem *sonst* – in manchen Systematiken (so von Polenz 1985: 270; Fritsche 1981: 72; Jude¹³1971: 202 und 252; vgl. weitere Literaturhinweise bei Denissova 2001: 139, insbes. auch Fn. 10) zu den Disjunktiven gerechnet werden. Die sprachlichen Fakten werden aber besser damit eingefangen, dass diese Konnektoren semantisch eben als negativ-konditionale gefasst werden. Schon syntaktisch würde eine Aufnahme unter die disjunktiven die große Homogenität jener Klasse stören.

¹³ Eine deutlich niedrigere Frequenz im Vergleich zu anderen Varianten ist generell nur ein Indiz, dass eine eingeschränkte Akzeptabilität vorliegen *kann*, und bedeutet nicht zwangsläufig, dass ein Ausdruck nicht standardsprachlich ist (das komplizierte Verhältnis zwischen Frequenz und Akzeptabilität problematisiert z. B. Conrad 2010: 237).

Abb. 1 – Absolute Frequenzen im DeReKo



Außerdem deutet auch die unterschiedliche Erwähnung in Grammatiken und Wörterbüchern darauf hin: Die Formen *ansonst* oder *sonsten* werden noch nicht einmal bei Pasch et al. (2003) gelistet und tauchen auch in modernen Wörterbüchern nur hier und da auf (*ansonst* etwa im DUW, 2011, aber nicht im HDG, 1984), oder sie werden zwar genannt, aber es wird nicht allen ein Lemma gewidmet: So hat Duden (21989: 682) als eigenes Lemma nur *sonst*.

Um die Unterschiede in ihrem Gebrauch näher zu bestimmen, müssen jetzt potenzielle „außersprachliche“ Variationsfaktoren wie Zeit, Raum oder Textsorte/Register untersucht und auch die Frequenzen der einzelnen Verwendungen der Lexeme unter die Lupe genommen werden.

3 Frequenz und Varianz der Lexeme

Untersucht wird in diesem Abschnitt die Varianz der Lexemfrequenzen im Gesamtkorpus. Das heißt vor allem auch, dass die Lexeme hinsichtlich ihrer Verteilung auf Teilkorpora miteinander verglichen werden. Das Gesamtkorpus ist gegliedert im Hinblick auf verschiedene Parameter,¹⁴ von denen sich bei der Variation der vier Konnektoren folgende als besonders relevant erwiesen:

- Staat (Teilkorpora: Deutschland, Österreich, Schweiz)
- Region (Teilkorpora: Nordwest, Nordost, Mittelwest, Mittelost, Mittelsüd, Südwest, Südost)
- Register (Teilkorpora: Publikumspresse, sonstige Gebrauchstexte, literarische Texte)

¹⁴ Vgl. Abschnitt 1; Genaueres zur Unterteilung des Gesamtkorpus, zu Parametern, Motivierung der für die Teilkorpora relevanten Kategorien, Korpusgrößen etc. in Konopka (2012, <http://hypermedia.ids-mannheim.de/call/public/korpus.ansicht?v_id=4755> bis <...id=4760>).

Die Abhängigkeit der Lexemfrequenzen vom Faktor Zeit wird unabhängig von obiger Einteilung in (virtuelle) Teilkorpora geprüft: Nahezu alle Texte des Gesamtkorpus sind mit einer Jahresdatierung versehen, was eine flexible Modellierung der Zeitabschnitte ermöglicht; außerdem ist beim Faktor Zeit ein Vergleich der Frequenzen im Untersuchungskorpus mit Frequenzen in Korpora zu älteren Sprachstufen bzw. in stärker historisch ausgerichteten Korpora möglich. Da die vier Konnektoren grundsätzlich unterschiedliche historische Entwicklungen zurückgelegt haben, soll hier mit der Betrachtung des Faktors Zeit begonnen werden.

3.1 Zeit

Um die Abhängigkeit der Lexemfrequenzen von verschiedenen Faktoren im Untersuchungskorpus zu berechnen und zu visualisieren, werden in R erzeugte Chi-Quadrat-basierte Assoziationsplots verwendet. Sie stellen die standardisierten Residuen dar und basieren auf zweidimensionalen Kontingenztafeln, sodass ein direkter Vergleich zwischen den Lexemen möglich ist. In Abbildung 3 wie auch in den weiteren Abbildungen mit Assoziationsplots sind die Rechtecke für die Lexeme relativ zu einer Grundlinie positioniert, die die Unabhängigkeit (vom Faktor) repräsentiert. Die Rechtecke befinden sich über der Grundlinie, wenn die beobachtete Frequenz höher ist als die erwartete Frequenz, und sie befinden sich unter der Linie, wenn die beobachtete Frequenz niedriger ausfällt als die erwartete. Die Höhe der Rechtecke korrespondiert mit dem Ausmaß der Abweichung von der Erwartung, ihre Breite mit der erwarteten Frequenz der Korpusbelege. Die Intensität der Schattierung schließlich zeigt die Signifikanz der Abweichung an.

Abbildung 2 ist insbesondere im Hinblick auf die häufigen Konnektoren, *ansonsten* und *sonst*, interessant, weil sie bei diesen komplementäre Entwicklungen zeigt. Die Frequenz von *ansonsten* nimmt in Relation zur Frequenz von *sonst* im Untersuchungskorpus mehr oder weniger kontinuierlich zu. Der Wendepunkt liegt dabei zwischen 2000 und 2010. In den Texten nämlich, die mit 2010 und später datiert sind, fällt die beobachtete Frequenz von *ansonsten* zum ersten Mal höher aus als die erwartete, während es sich bei *sonst* umgekehrt verhält. Bei den beiden übrigen Konnektoren *ansonst* und *sonsten* werden keine vergleichbaren Regularitäten sichtbar. Lediglich ihre Randständigkeit im Sprachgebrauch wird hier gespiegelt.

Die Rivalität zwischen *ansonsten* und *sonst* lässt sich im DWDS-Kernkorpus, das nach Dekaden und Textsorten ausgewogen das 20. Jh. dokumentieren soll,¹⁵ noch weiter zurückverfolgen, vgl. Abbildung 3 und 4.

¹⁵ Vgl. < <http://www.dwds.de/ressourcen/kernkorpus/> >

Abb. 2 – Konnektorenfrequenz in Abhängigkeit vom Jahrzehnt

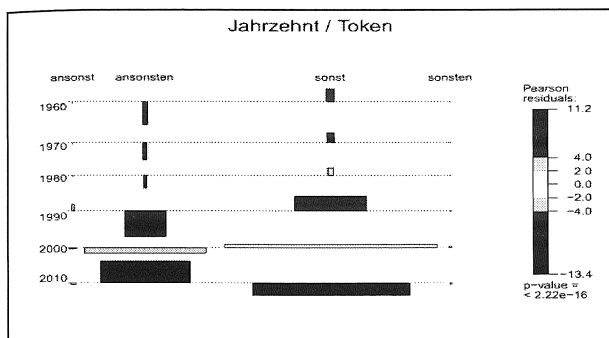


Abb. 3 – Wortverlauf (pro 1 Mio. Token) für *ansonsten* (nur Kleinschreibung) im DWDS-Kernkorpus (http://www.dwds.de/?qu=ansonsten&submit_button=Suche&view=4 am 20.06.2013)

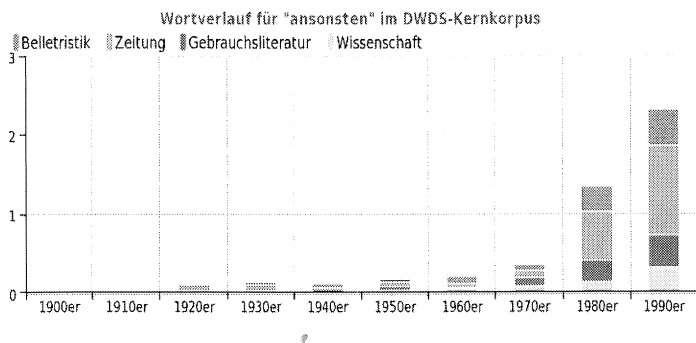
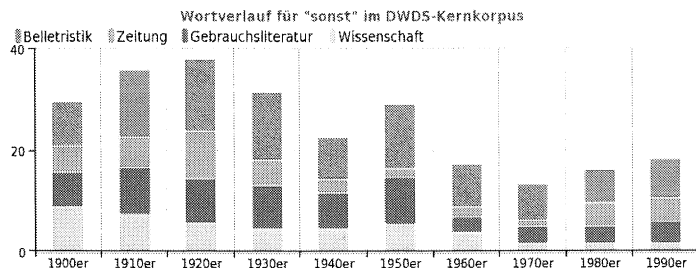


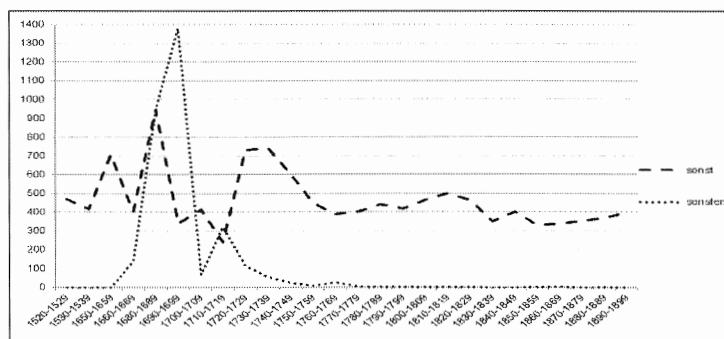
Abb. 4 – Wortverlauf (pro 1 Mio. Token) für *sonst* (nur Kleinschreibung) im DWDS-Kernkorpus (http://www.dwds.de/?qu=ansonsten&submit_button=Suche&view=4 am 20.06.2013)



Auffällig ist, dass *ansonsten* in den Texten aus dem Anfang des 20. Jahrhunderts praktisch nicht vorkommt und erst seit den sechziger Jahren als ernsthafter Konkurrent für *sonst* betrachtet werden kann.

Die beiden übrigen Konnektoren spielen im DWDS-Kernkorpus so gut wie keine Rolle,¹⁶ sodass für sie keine Wortverläufe verfügbar sind. Aber es lohnt sich, noch weiter zurückzublicken: Im Archiv HIST von COSMAS II,¹⁷ das vor allem Texte aus der Zeit vom 16. Jh. bis zur ersten Hälfte des 20. Jhs. mit einem Schwerpunkt auf dem 19. Jh. erfasst,¹⁸ sind beide mit *an-* präfigierten Varianten marginal belegt – *ansonst* mit 10 (1816–1901) und *ansonsten* mit 7 Treffern (1846–1920). Dagegen ist *sonsten* mit 414 Treffern (1662–1919) reichlich vertreten. *Sonsten* ist in den Texten aus der Zeit von etwa 1680 bis 1730 – ähnlich wie *ansonsten* in den Gegenwartstexten – sogar ein ernstzunehmender Konkurrent für *sonst* und zeigt in dieser Periode eine komplementäre Frequenzentwicklung (vgl. Abbildung 5).¹⁹

Abb. 5 – Wortverläufe (pro 1 Mio. Token) für *sonst* und *sonsten* im Archiv HIST von COSMAS II



¹⁶ Zu *ansonst/Ansonst* findet man 51 Treffer, zu *sonsten/Sonsten* keinen einzigen. Dies mag an der Korpusgröße liegen (100 Mio. Token) – zur Erinnerung: Unser Untersuchungskorpus umfasst über 4 Mrd. Token.

¹⁷ Ein am Institut für Deutsche Sprache in Mannheim entwickeltes Korpusrecherche- und -analysesystem, vgl. <<http://www.ids-mannheim.de/cosmas2/>>.

¹⁸ Über 98% der 65,2 Mio. Wörter des Archiv HIST gehören Texten an, die auf die Zeit vor 1920 datiert sind (Stand 10. 06. 2013).

¹⁹ Es darf nicht unerwähnt bleiben, dass die Frequenzen pro 1 Mio. Token im Archiv HIST insgesamt viel höher ausfallen als im DWDS-Kernkorpus, vgl. – für *sonst* – ca. 400 Vorkommen im Archiv HIST 1890–1899 versus ca. 30 Vorkommen im DWDS-Korpus 1900–1910. Die Frequenzen pro 1 Mio. Token sind im DWDS-Kernkorpus offensichtlich anders berechnet worden. Allerdings wurden oben die Verläufe für *ansonsten* und *sonst* bzw. für *sonst* und *sonsten* in jeweils einem Korpus/Archiv miteinander verglichen, sodass die obige Gesamtargumentation aufrechterhalten werden kann.

Die Gesamtentwicklung einzuschätzen, fällt leichter, wenn man sprachreflexive Werke konsultiert: Adelung (1780: 527) merkt bei *sonst* an: „in den gemeinen Mundarten auch *sonsten*“. *Ansonst/ansonsten* wird in einer späteren Auflage seines Wörterbuchs dem Oberdeutschen (Adelung 1808b: 150), *ansonst* allein auch dem Kanzleistil zugeordnet (Adelung, 1808a: 372); die Zuordnung zum Oberdeutschen/Süddeutschen wird öfter wiederholt (z. B. Paul 1992: 44). *Sonsten* wird demgegenüber viel später bei Paul (2002: 80) als „schriftsprachlich“ eingeschätzt.

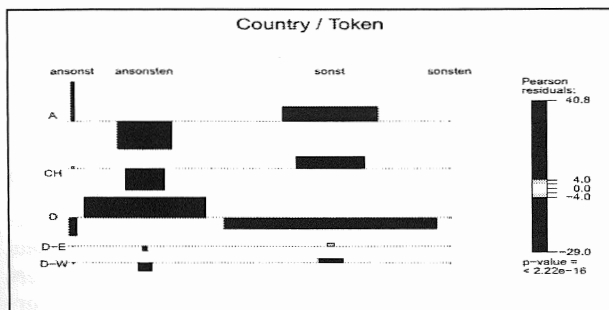
Es entsteht ein Gesamtbild: Die ursprüngliche Form *sonst* bildet eine Nebenform *sonsten* heraus, die Ende des 17. und in der ersten Hälfte des 18. Jahrhunderts auch schriftsprachlich zu einer ernstzunehmenden Konkurrenz wird. Beide Varianten werden durch die Erweiterung um *an-* erfasst, die sich aber interessanterweise nur bei der Nebenform *sonsten* langfristig durchsetzen kann. Bei Adelung und Paul schimmern in Bezug auf die historische Rivalität zwischen *sonst* und *ansonst*, *sonsten* und *ansonsten* Medium (in „gemeine Mundart“, „schriftsprachlich“), Register (in „Kanzleistil“) und Region (in „Oberdeutsch“) als Einflussfaktoren durch. Ob und welche Rolle solche Faktoren in der heutigen Variation der vier Konnektoren spielen, klärt der folgende Abschnitt.

3.2 Raum und Register

Ähnlich wie aus der gegenwartsnahen zeitlichen Perspektive verhalten sich die häufigen Konnektoren *sonst* und *ansonsten* auch in Bezug auf den Raum und das Register komplementär.

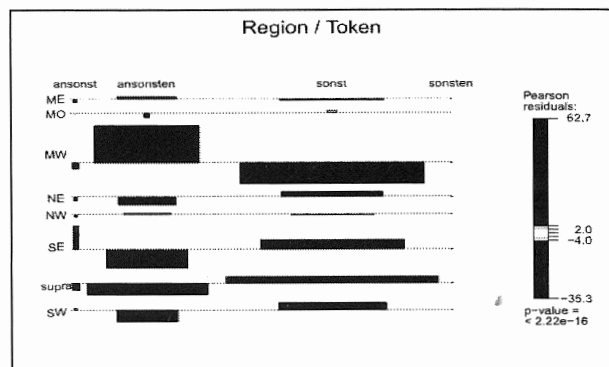
Während die Häufigkeit von *sonst* in österreichischen und schweizerischen Texten deutlich die Erwartung übersteigt, bleibt *ansonsten* in diesen Texten hinter der Erwartung zurück (vgl. Abbildung 6). Dagegen in bundesdeutschen Texten dreht sich dieses Verhältnis um. Aus der nationalsprachlichen Perspektive ist auch *ansonst* interessant, das besonders häufig in österreichischen Texten erscheint.

Abb. 6 – Konnektorenfrequenz in Abhängigkeit vom Faktor Staat



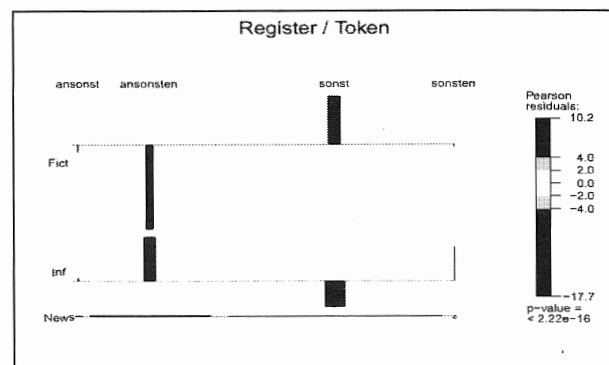
Steigt man auf die Ebene der Großregionen hinab, wird deutlich, dass den größten Anteil an der Dominanz von *ansonsten* in Deutschland Texte aus dem Mittelwesten haben (vgl. Abbildung 7).

Abb. 7 – Konnektorenfrequenz in Abhängigkeit vom Faktor Region



Zieht man den Faktor Register heran, wird deutlich (siehe Abbildung 8), dass *sonst* weit häufiger als erwartet in literarischen Texten erscheint, *ansonsten* dagegen in Gebrauchstexten (ohne journalistische Texte).

Abb. 8 – Konnektorenfrequenz in Abhängigkeit vom Faktor Register



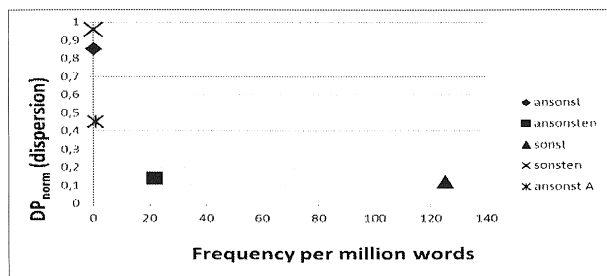
3.3 Streuung

Man kann annehmen, dass standardsprachliche Varianten häufiger auftreten und gleichmäßiger über die Teilkorpora verteilt sein müssen als Varianten, deren Standardsprachlichkeit in Frage steht. In Abbildung 9 werden kombiniert die Streuung der vier Konnektoren und ihre relative Frequenz (pro 1 Mio. Wör-

ter) dargestellt. Für die Streuung wird das Maß DP_{norm} von Gries verwendet, das Werte zwischen 0 und 1 einnehmen kann (je höher die DP_{norm} , desto ungleichmäßiger die Verteilung) und speziell für Teilkorpora unterschiedlicher Größe entworfen wurde.

Abbildung 9 zeigt, dass *ansonsten* und *sonst* trotz eines großen Frequenzunterschieds annähernd gleich verteilt sind. Unserer Intuition nach gehören sie heute gleichermaßen zum Standard, und aus dieser Perspektive erscheint DP_{norm} als ein deutlich besserer Indikator der Standardzugehörigkeit als die relative Frequenz. Was die beiden selteneren Konnektoren angeht, ist *ansonst* sowohl häufiger (vgl. auch Abb. 1 weiter oben) als auch gleichmäßiger verteilt, wenn auch von den Werten für *sonst* und *ansonsten* noch weit entfernt. In den österreichischen Texten ist es etwas häufiger und vor allem deutlich gleichmäßiger verteilt (vgl. *ansonst A*).

Abb. 9 – Streuung und relative Frequenz im Gesamtkorpus



Das Zwischenfazit lautet also: Legt man die am Anfang dieses Abschnitts geäußerte Annahme zugrunde, so bestätigt unser statistisches Instrumentarium und vor allem die DP_{norm} , dass

- *sonst* und *ansonsten* standardsprachliche Varianten sind und
- *ansonst* und *sonsten* nicht (mehr) zum allgemeinen Standard gehören.

Es konnten auch einige zeitliche, räumliche und registerbezogene Charakteristika der vier Varianten spezifiziert werden. Da sich das Instrumentarium bewährt hat, ist auch der zusätzliche Hinweis, den uns die verwendeten Maße geben, ernst zu nehmen, und zwar dass *ansonst* einem Teilstandard im Südwesten nahe stehen könnte. Dies wird noch genauer zu untersuchen sein.

4 Frequenz und Varianz der syntaktischen Verwendungen

Die Untersuchung der syntaktischen Verwendungen der Konnektoren erfordert eine sorgfältige qualitative Auswertung der Korpusbelege, sodass deren Menge auf drei zufällig zusammengestellte Stichproben von je 200 Belegen für

sonst, ansonsten und ansonst zuzüglich **aller** Belege für *sonsten* (58) beschränkt werden musste. In diesem *Stichprobenkorpus* konnten folgende syntaktische Verwendungen der vier Konnektoren unterschieden werden (zu Beispielen vgl. Abschnitt 2):

- Adverbkonnektor im Vorfeld (VF),
- Adverbkonnektor im Mittelfeld (MF),
- Phraseninterner Adverbkonnektor (z. B. mit Adjektivbezug, XP),
- Postponierer (PostP).

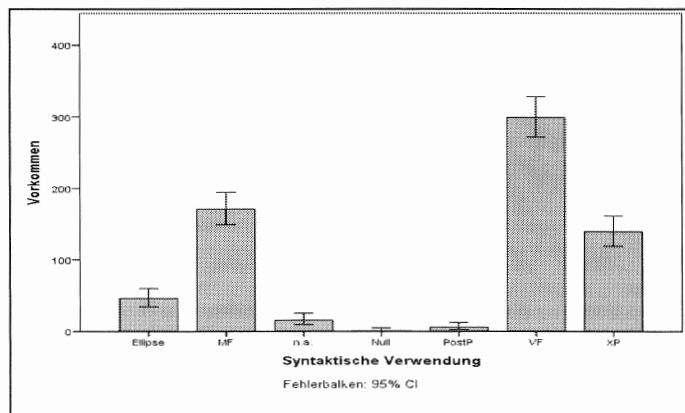
Hinzu kommen elliptische Verwendungen (*Ellipse*), die sich meist als Adverbkonnektoren klassifizieren, jedoch nicht eindeutig hinsichtlich des Satzfeldes (VF/MF) spezifizieren lassen, z. B.:

- (14) Silverstone: Clueless – Was **sonst**! (Ressort: Chronik/Wien-Journal; Die Presse, 30. 12. 1995)

Schließlich kamen vereinzelt Verwendungen als Adverbkonnektor in der Nullposition vor (*Null/0*, vgl. Fußnote 10).

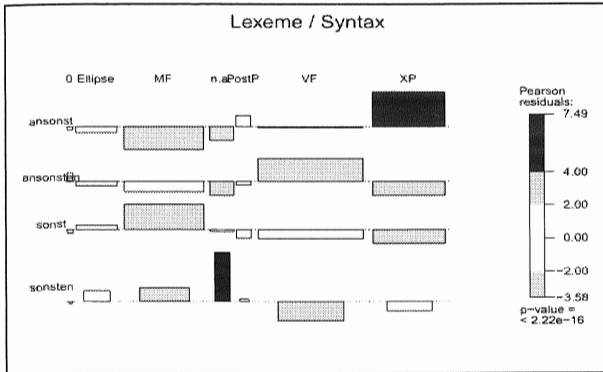
Zwischen den verschiedenen Verwendungen gibt es große Frequenzunterschiede. Sie werden ohne Rücksicht auf das jeweilige Lexem in Abbildung 10 veranschaulicht. Die Fehlerbalken zeigen die Konfidenzintervalle bei einem Konfidenzkoeffizient von 95%. Die klassischen Adverbkonnektor-Verwendungen (VF, MF, XP) erweisen sich darin als deutlich häufiger als andere. Die Verwendung als Postponierer hingegen nimmt hinsichtlich der Frequenz die vorletzte Stelle ein und erscheint als marginal.

Abb. 10 – Frequenz syntaktischer Verwendungen im Stichprobenkorpus



In Abbildung 11 wird die Verteilung der Verwendungen auf die einzelnen Lexeme visualisiert.

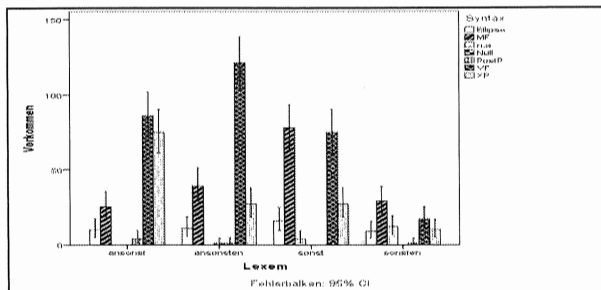
Abb. 11 – Frequenz syntaktischer Verwendungen in Abhängigkeit vom Konnektor



Es zeigt sich einerseits, dass es Korrelationen zwischen den Verwendungen und bestimmten Konnektoren gibt. Besonders signifikant ist die Abweichung der beobachteten Frequenz von der erwarteten für die Verwendung als phraseninterner Adverbkonnektor (*XP*) bei *ansonst*. Aber auch die Verwendung als Adverbkonnektor im Vorfeld (*VF*) bei *ansonsten*, die Verwendung als Adverbkonnektor im Mittelfeld (*MF*) bei *sonst* und *sonsten* und die Verwendung als Ellipse bei *sonsten* übersteigen deutlich die Erwartung.

Andererseits wird deutlich, dass die Konnektoren jeweils bevorzugte Verwendungen haben: *Ansonsten* präferiert stark die Verwendung als Adverbkonnektor im Vorfeld, *ansonst* die als Adverbkonnektor im Vorfeld und als phraseninterner Adverbkonnektor und *sonst* die als Adverbkonnektor im Mittelfeld oder Vorfeld (vgl. Abbildung 12).

Abb. 12 – Frequenz syntaktischer Verwendungen im Stichprobenkorpus je Konnektor

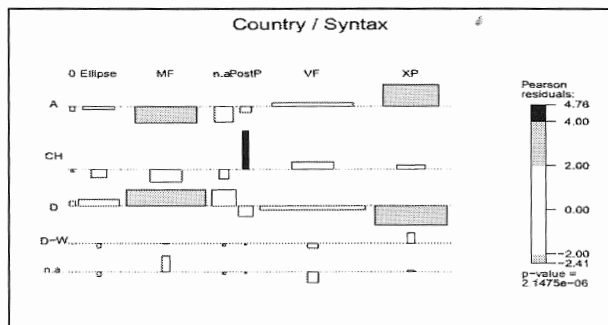


Um Missverständnissen vorzubeugen: *Sonst* ist aufgrund seiner im Vergleich zu anderen Konnektoren sehr hohen Frequenz (vgl. Abb. 1 und 9 weiter oben) dennoch die erste Wahl für alle syntaktischen Verwendungen mit Ausnahme der Postponierer-Verwendung. Neben unterschiedlichen Korrelationen der syntaktischen Verwendungen und der einzelnen Konnektoren lassen sich bei den syntaktischen Verwendungen vor allem räumliche Besonderheiten feststellen.

4.1 Raum

Untersucht man die räumliche Verteilung der syntaktischen Verwendungen im *Stichprobenkorpus*, fällt auf, dass bestimmte Verwendungen regional über- bzw. unterrepräsentiert erscheinen (vgl. Abbildung 13).

Abb. 13 – Frequenz syntaktischer Verwendungen in Abhängigkeit vom Faktor Staat



Signifikante Abweichungen von der erwarteten Frequenz betreffen drei Verwendungen:

- Die Verwendung als Adverbkonnektor im Mittelfeld (*MF*) ist überrepräsentiert in bundesdeutschen und unterrepräsentiert in österreichischen Texten.
- Die Verwendung als Postponierer ist überrepräsentiert in den schweizerischen Texten.
- Die Verwendung als phraseninterner Adverbkonnektor (*XP*) ist überrepräsentiert in österreichischen und unterrepräsentiert in bundesdeutschen Texten.

Befassen wir uns mit den beiden letzteren Auffälligkeiten in der Hoffnung, dass sie uns helfen, den Status der beiden selteneren Konnektoren *ansonst* und *sonsten* besser einzuschätzen.

Die Postponierer-Verwendung verteilt sich im Stichprobenkorpus nur auf die Konnektoren *ansonst*, *ansonsten* und *sonsten*, und zwar im Verhältnis

4 : 1 : 1.²⁰ Die niedrige Anzahl der Belege lässt es jedoch nicht zu, eine statistisch signifikante Korrelation zwischen *ansonst* und Postponierer-Verwendung zu postulieren.

Die Verwendung als phraseninterner Konnektor wiederum verteilt sich auf *ansonst*, *ansonsten*, *sonst* und *sonsten*, und zwar im Verhältnis 75 : 27 : 27 : 6. Eine besondere Affinität von *ansonst* zu dieser Verwendung stand ohnehin schon fest (vgl. Abbildung 11 weiter oben), wie auch übrigens die der österreichischen Texte zu *ansonst* (vgl. Abbildung 6). Bei der Verwendung als phraseninterner Konnektor entstammen österreichischen Texten 57 der 75 *ansonst*-Belege, und so verdichten sich die Hinweise darauf, dass *ansonst* in der Verwendung als phraseninterner Konnektor in österreichischen Texten bzw. in Texten aus dem Südosten akzeptabler ist als in Texten aus anderen Regionen, ja vielleicht sogar dass *ansonst* in dieser speziellen Verwendung im Südosten einem regional definierten Standard nahesteht.

Eine weitergehende Klärung können hier Untersuchungen zur Streuung von *ansonst* als phraseninterner Adverbkonnektor bringen.

4.2 Streuung von *ansonst*

In der Stichprobe von 200 *ansonst*-Belegen konnte sich als phrasenintern gewertetes *ansonst* auf Verschiedenes beziehen, z. B. auf

- eine Adjektivphrase, vgl.:
(15) Das Image der *ansonst* so korrekten Deutschen ist beschädigt. (Nürnberger Nachrichten, 29. 1. 2005, o. S.)
- eine Präpositionalphrase, vgl.:
(16) Für die Mütter galt gleiches wie *ansonst* bei ihren Töchtern: Ein Punktesystem, bei dem Schlamperei und Fehler geahndet werden. (Neue Kronen-Zeitung, 17. 02. 1996, S. 18)
- ein Pronomen, vgl.:
(17) Es gibt einen Kader, *ansonst* nichts. (Neue Kronen-Zeitung, 25. 05. 1997, S. 58)

Dabei machten die Fälle mit Bezug auf eine Adjektivphrase 71 von 75 einschlägigen Belegen aus. In 70 Fällen stand dabei der Konnektor direkt hinter einem Artikel (wie in (15) oder (6) weiter oben), und ein solches Distributionsmuster ist in einem Korpus maschinell gut auffindbar – im Gegensatz zu allen anderen Distributionsmustern der phraseninternen Verwendung, die als schwer prognostizierbar erscheinen. So wurde nach dem Muster²¹

²⁰ Zur Erinnerung: Pro Konnektor sind im Stichprobenkorpus 200 Belegsätze enthalten, außer bei *sonsten* (nur 58 Belegsätze).

²¹ Berücksichtigt wurden nur definite, indefinite und possessive Artikel, um die Gefahr zu

der/die/das/des/dem/den/

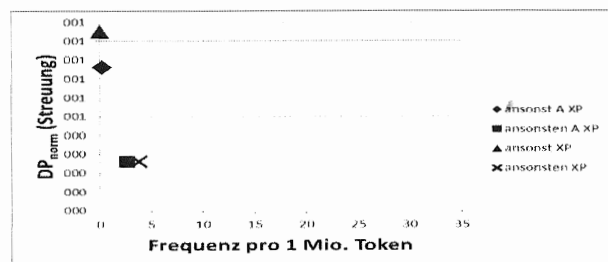
ein/eine/eines/einer/einem/einen

+ Konnektor

mein/meine/meines/meiner/meinem/meinen

in allen österreichischen Presstexten des auf DeReKo basierenden Archivs W von COSMAS II gesucht, und zwar für *ansonst* und – zum Vergleich – auch für *ansonsten*. Abbildung 14 fasst die DP_{norm} -Werte und die relativen Häufigkeiten, die sich dabei ergaben, zusammen (gekennzeichnet mit A XP) und vergleicht sie mit den Werten im gesamten Archiv W.

Abb. 14 – Streuung und relative Frequenz von *ansonst* und *ansonsten* in phraseninterner Verwendung



Es zeichnet sich folgendes Bild ab: *Ansonst* als phraseninterner Adverbkonnektor erscheint in österreichischen Texten zwar nur unwesentlich häufiger als in Gesamt-DeReKo, es ist in diesen aber deutlich gleichmäßiger verteilt. Nichtsdestoweniger ist *ansonsten* in österreichischen Texten wie auch in Gesamt-DeReKo immer noch deutlich besser gestreut.

Das Zwischenfazit der Untersuchung der syntaktischen Verwendungen lautet also:

- Die Verwendungen als Adverbkonnektor im Vorfeld und Mittelfeld sind am häufigsten. In beiden sind zumindest *sonst* und *ansonsten* standardsprachlich, dabei zeigt *ansonsten* eine starke Präferenz für das Vorfeld.
- Die Verwendung als phraseninterner Adverbkonnektor ist insgesamt gesehen fast genauso üblich wie die beiden oben genannten Verwendungen. Sie ist überrepräsentiert in österreichischen Texten. In diesen ist *ansonst* als phraseninterner Adverbkonnektor zu den üblichen *ansonsten* und *sonst* auch beinahe konkurrenzfähig.

minimieren, dass in der gleichen Position andere Wortarten mit identischen Formen erscheinen. Um die mit den definiten Artikeln formal identischen Relativpronomina auszuschließen, wurden Belege ausgeschlossen, in denen vor dem Muster ein Komma steht.

- Die Postponierer-Verwendung gehört offensichtlich nicht der Standardsprache und auch nicht einem großregional definierbaren Teilstandard an. Allerdings kommt sie häufiger in schweizerischen Texten vor.

5 Folgerungen

5.1 Linguistische Phänomene

Die Behauptung, dass *sonst*, *ansonsten*, *ansonst* und *sonsten* Varianten sind, darf nicht unkommentiert stehen bleiben. Die Analysen zeigen, dass die vier Konnektoren

- unterschiedlich frequent sind,
- sich unterschiedlich auf das Gesamtkorpus verteilen und
- unterschiedliche syntaktische Verwendungen präferieren.

Nur *sonst* und *ansonsten* gehören ohne Einschränkungen der Standardsprache an, was ihre Streuungswerte (DP_{norm}) spiegeln, wobei die Verwendung als Postponierer heute im Allgemeinen nicht standardsprachlich erscheint. In unseren Daten begegnet sie vor allem in schweizerischen Texten. Sie wird in erster Linie mit dem insgesamt randständigen *ansonst* realisiert.

Es wird deutlich, dass sich Einschätzungen hinsichtlich der Standardsprachlichkeit der Varianten immer auf Kombinationen aus einem Lexem und einer bestimmten (manchmal ziemlich spezifischen) Verwendung beziehen und dass dabei u. U. außersprachliche Parameter berücksichtigt werden müssen. So legen unsere Varianz- und Streuungsdaten nahe, dass der *ansonsten* nichtstandardsprachliche Konnektor *ansonst* in österreichischen Texten akzeptabler erscheint, wenn er phrasenintern gebraucht wird. Die Einschätzungen zur Standardsprachlichkeit aller Kombinationen werden in Tabelle 2 zusammengefasst. Die Schattierung signalisiert darin die Standardnähe bzw. -ferne: Die nicht schattierten Kombinationen gehören ohne Einschränkungen dem (allgemeinen) Standard an, die dunkelgrau schattierten liegen am anderen Pol der Skala. Dazwischen liegt das hellgrau schattierte *ansonst* in der phraseninternen Verwendung in österreichischen Texten.

Tab. 2 – Standardsprachlichkeit der Konnektoren und ihrer Verwendungen I

Syntax/Konnektor	<i>sonst</i> (allgemeiner Standard)	<i>ansonsten</i> (allgemeiner Standard)	<i>ansonst</i> (vor allem in Österreich)	<i>sonsten</i> (Nonstandard)
Adverbkonnektor MF/VF	allgemeiner Standard	allgemeiner Standard	Nonstandard	Nonstandard
Adverbkonnektor XP	allgemeiner Standard	allgemeiner Standard	standardnah? (in Österreich)	Nonstandard
Postponierer (vor allem in der Schweiz)	Nonstandard	Nonstandard	Nonstandard	Nonstandard

Wir sind überzeugt, dass der Übergang von den uneingeschränkt nicht-standardsprachlichen Kombinationen zu denen, deren Standardsprachlichkeit absolut nicht in Frage steht, mit noch mehr graduellen Zwischenstufen modelliert werden muss, was in Tabelle 3 veranschaulicht wird.

Tab. 3 – Standardsprachlichkeit der Konnektoren und ihrer Verwendungen II

Syntax/Konnektor	<i>sonst</i> (allgemeiner Standard)	<i>ansonsten</i> (allgemeiner Standard)	<i>ansonst</i> (vor allem in Österreich)	<i>sonsten</i> (Nonstandard)
Adverbkonnektor im Mittelfeld	allgemeiner Standard	allgemeiner Standard	Nonstandard	Nonstandard
Adverbkonnektor im Vorfeld	allgemeiner Standard	allgemeiner Standard	Nonstandard	Nonstandard
Adverbkonnektor XP	allgemeiner Standard	allgemeiner Standard	standardnah? (in Österreich)	Nonstandard
Postponierer (vor allem in der Schweiz)	Nonstandard	Nonstandard	Nonstandard	Nonstandard

An einem Ende des Spektrums befindet sich *sonst* in der Postponierer-Verwendung, das absolut nicht-standardsprachlich, wenn nicht überhaupt inakzeptabel erscheint und in der Tabelle dunkelgrau schattiert ist. An dem anderen Ende stehen *ansonsten* und *sonst* jeweils in der Verwendung, zu der sie die größte Affinität zeigen; sie sind in der Tabelle nicht unterlegt. Dazwischen liegen weitere Kombinationen aus Lexemen und Verwendungen, deren Standardnähe bzw. -ferne durch verschiedene Grautöne veranschaulicht wird.

5.2 Methodologie

In einem ersten Schritt wurden die grammatischen Phänomene traditionell linguistisch analysiert. Danach wurden ihre Frequenz und Varianz in Korpora und ihren Teilen untersucht, wobei zunächst Lexeme und danach – von Lexemen zunächst unabhängig – syntaktische Verwendungen fokussiert wurden. Dies sollte differenzierte Hinweise darauf geben, ob die Phänomene (1) der Standardsprache oder (2) einem regional oder auch anders definierten spezifischen Standard angehören bzw. (3) ob sie außerhalb des Standards stehen. Die entstandenen Hypothesen über spezifische Kombinationen aus Lexemen, syntaktischen Verwendungen und Ausprägungen außersprachlicher Parameter wurden im nächsten Schritt validiert. Aufgrund von Vergleichen der Frequenz- und Streuungswerte für verschiedene Faktorenkombinationen wurden Einschätzungen ihrer Standardsprachlichkeit vorgenommen.

Die Untersuchung beruhte auf Recherchen in großen, nach außersprachlichen Faktoren parametrisierten Korpora und auf statistischen Analysen der

Frequenz und Streuung der Phänomene. Mit den Chi-Quadrat-basierten Assoziationsplots konnte die Verteilung der Phänomene auf virtuelle Teilkorpora visualisiert werden, die verschiedenen Ausprägungen eines mutmaßlichen Variationsfaktors entsprachen wie Zeit, Raum und Register. Auf diese Weise konnte die Varianz der Phänomene zum Teil überhaupt erst entdeckt werden. Vor allem aber konnten so Hypothesen über spezifische Abhängigkeiten der Phänomene von den Faktoren gebildet werden, die im Weiteren verfolgt wurden. Dabei half die DP_{norm} von Gries, ein speziell für Korpora verschiedener Größe entworfenen Streuungsmaß, zu erfassen, wie gleichmäßig die Phänomene innerhalb des Gesamtkorpus bzw. innerhalb eines Teilkorpus verteilt sind. Im Hinblick auf den Faktor Zeit wurde auch auf das Mittel des Vergleichs mit historisch gesehen anders ausgerichteten Korpora zurückgegriffen. All dies half, die Affinität der Phänomene zur Standardsprache weit präziser einzuschätzen, als dies auf der Grundlage der reinen Intuition und der Betrachtung von bloßen Frequenzen im Korpus möglich ist, und erlaubte schließlich eine korpuslinguistisch fundierte Modellierung des Kontinuums zwischen Standard und Nonstandard im Bereich der negativ-konditionalen Konnektoren.

5.3 Ausblick

Das Projekt „Korpusgrammatik“ ist als „work in progress“ zu betrachten, und es zielt methodologisch nicht zuletzt darauf ab, eine transparente quantitativ-statistische Basis für die Validierung von linguistischen Hypothesen zu schaffen. Die hier verwendeten korpuslinguistischen Methoden und statistischen Instrumente werden in den nächsten Studien weiter evaluiert. Unter anderem werden dabei Instrumente zur Erfassung der Streuung eines Phänomens im Korpus und in seinen Teilen wie die hier verwendete DP_{norm} weiter zu untersuchen sein. Dabei muss insbesondere den Fragen nachgegangen werden, wie der Übergang vom Nonstandard zum Standard auf die Skala der Streuungswerte genau abzubilden ist und wie stark die Streuungswerte von Frequenzwerten abhängen. Um die Entdeckung von Variations- und Varianzfällen zu optimieren und eine detailliertere Untersuchung der Variationsfaktoren zu ermöglichen, werden schließlich auch andere korpuslinguistische Methoden in die Projektarbeit integriert, wie etwa maschinelles Lernen zur Vorhersage der Variantenwahl.²²

LITERATUR

ADELUNG, J. C. (1780): *Versuch eines vollständigen grammatisch-kritischen Wörterbuches Der Hochdeutschen Mundart: mit beständiger Vergleichung der übrigen Mundarten, besonders aber der Oberdeutschen*. Vierter Theil, von Sche – V. Leipzig: Breitkopf.

²² Vgl. Bubenhofer et al. (2012).

ADELUNG, J. C. (1808a): *Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart: mit beständiger Vergleichung der übrigen Mundarten, besonders aber des Oberdeutschen*. Erster Theil, von A – E. Wien: Pichler.

ADELUNG, J. C. (1808b): *Grammatisch-kritisches Wörterbuch der hochdeutschen Mundart: mit beständiger Vergleichung der übrigen Mundarten, besonders aber des Oberdeutschen*. Vierter Theil, von Seb – Z. Wien: Pichler.

BIBER, D. et al. (1999; 2006): *Longman Grammar of Spoken and Written English*. Harlow: Longman.

BREINDL, E. – VOLODINA, A. – WAßNER, U. H. (im Ersch.). *Handbuch der deutschen Konnektoren. Teil 2: Semantik*. Berlin / New York: de Gruyter (Schriften des Instituts für Deutsche Sprache).

BUBENHOFER, N. (2012): Statistische Methoden für korpusgrammatische Analysen. In: *Grammis 2.0, Korpusgrammatik*. Mannheim: Institut für Deutsche Sprache. <http://hypermedia.ids-mannheim.de/call/public/korpus.ansicht?v_id=4683> am 27. 06. 2013.

BUBENHOFER, N. – HEIN, K. – BRINKMANN, C. (2012): Maschinelles Lernen zur Vorhersage von Fugenelementen in nominalen Komposita. In: *Grammis 2.0, Korpusgrammatik*. Mannheim: Institut für Deutsche Sprache. <http://hypermedia.ids-mannheim.de/call/public/korpus.ansicht?v_id=4697> am 27. 06. 2013.

COHEN, A. (1980): *On the graphical display of the significant components in a two-way contingency table*. In: *Communications in Statistics – Theory and Methods*, A9 (10), 1025–1041.

CONRAD, S. (2010): What can a corpus tell us about grammar? In: O’Keeffe, A., McCarthy, M. (eds.), *The Routledge Handbook of Corpus Linguistics*. London, New York: Routledge, 227–240.

COSMAS II. <<http://www.ids-mannheim.de/cosmas2/>> am 27. 06. 2013.

Deutsches Referenzkorpus (DeReKo). <<http://www.ids-mannheim.de/kl/projekte/korpora/>> am 27. 06. 2013.

DENISOVA, M. D. (2001): Parataxe versus Koordination (am Beispiel der *a mo-* bzw. *sonst/weil-V-2*-haltigen Satzkomplexe des gesprochenen Deutschen und Russischen). In: W. Gladrow, R. Hammel (Hrsg.), *Beiträge zu einer russisch-deutschen kontrastiven Grammatik*. Frankfurt a. M.: Lang, 133–150 (Berliner Slawistische Arbeiten 15).

DUDEN (21989): Band 7. *“Etymologie”: Herkunftswörterbuch der deutschen Sprache*. 2., völlig neu bearb. u. erw. Aufl. Mannheim: Dudenverlag.

DUW (2011): *Duden. Deutsches Universalwörterbuch: das umfassende Bedeutungswörterbuch der deutschen Gegenwartssprache*. 7., überarb. und erw. Aufl. Mannheim / Zürich: Dudenverlag.

DWDS-Kernkorpus. <<http://www.dwds.de/ressourcen/kernkorpus/>> am 27. 06. 2013.

FRIENDLY, M. (1992): *Graphical methods for categorical data*. SAS User Group International Conference Proceedings, 17, 190–200. <<http://www.math.yorku.ca/SCS/sugi/sugi17-paper.html>> am 27. 06. 2013.

FRITSCHKE, J. (1981): *Konnektivausdrücke, Konnektiveinheiten. Grundelemente der semantischen Struktur von Texten I*. Hamburg: Buske (Papiere zur Textlinguistik 30).

GRIES, S. T. (2008): Dispersions and adjusted frequencies in corpora. In: *International Journal of Corpus Linguistics* 13 (4), 403–437.

GRIES, S. T. (2009): Dispersions and adjusted frequencies in corpora: further explorations. In: *Language and Computers* 71 (1), 197–212.

HDG (1984): *Handwörterbuch der deutschen Gegenwartssprache: in zwei Bänden*. Berlin: Akademie-Verlag.

JUDE, W. K. (¹³1971): *Deutsche Grammatik*. 13. Aufl. Braunschweig: Westermann.

KONOPKA, M. (2012a): Gesamtkorpus. In: *Grammis 2.0, Korpusgrammatik*. Mannheim: Institut für Deutsche Sprache. <http://hypermedia.ids-mannheim.de/call/public/korpus.ansicht?v_id=4752> am 27. 06. 2013.

KONOPKA, M. (2012b): Variabilitätsfaktoren. In: *Grammis 2.0, Korpusgrammatik*. Mannheim: Institut für Deutsche Sprache. <http://hypermedia.ids-mannheim.de/call/public/korpus.ansicht?v_id=4755> am 27. 06. 2013.

KUBCZAK, J. – KONOPKA, M. (2008): Grammatical variation in Near-Standard German: a corpus-based project at the Institute for the German Language (IDS) in Mannheim. In: F. Štícha, M. Fried (eds.), *Grammar & corpus. Selected contributions from the conference Grammar and Corpora, Sept. 25 - 27, 2007, Liblice*. Prag: ACADEMIA.

MEYER, D. – ZEILEIS, A. – HORNIK, K. (2005): *The strucplot framework: Visualizing multi-way contingency tables with vcd*. Report 22, Department of Statistics and Mathematics, Wirtschaftsuniversität Wien (Research Report Series). <http://epub.wu.ac.at/dyn/openURL?id=oai:epub.wu-wien.ac.at:epub-wu-01_8a1> am 27. 06. 2013.

PASCH, R. – BRAUß, U. – BREINDL, E. – WAGNER, U. H. (2003): *Handbuch der deutschen Konnektoren. Linguistische Grundlagen der Beschreibung und syntaktische Merkmale der deutschen Satzverknüpfers (Konjunktionen, Satzadverbien und Partikeln)*. Berlin – New York: de Gruyter (Schriften des Instituts für Deutsche Sprache 9).

PAUL, H. (2002): *Deutsches Wörterbuch. Bedeutungsgeschichte und Aufbau unseres Wortschatzes*. 10., Aufl. von H. Henne, H. Kämper und G. Objartel. Tübingen: Niemeyer.

PAUL, H. (1992): *Deutsches Wörterbuch*. 9., Aufl. von H. Henne und G. Objartel unter Mitarbeit von H. Kämper-Jensen. Tübingen: Niemeyer.

VON POLENZ, P. (1985): *Deutsche Satzsemantik: Grundbegriffe d. Zwischen-den-Zeilen-Lesens*. Berlin: de Gruyter.